VOLUME 76 SEPARATE No. 11

ROCEEDINGS

AMERICAN SOCIETY
OF
CIVIL ENGINEERS

APRIL, 1950

AMERICAN SOCIETY OF CIVIL ENGINEERS FOUNDED 1852

Illincis U Library

LONG-TERM STORAGE CAPACITY OF RESERVOIRS

By H. E. Hurst

HYDRAULICS DIVISION

Headquarters of the Society
33 W. 39th St.
New York 18, N.Y.

PRICE $0.50 PER COPY

The Society is not responsible for any statement made or opinion expressed in its publications

---

Published at Prince and Lemon Streets, Lancaster, Pa., by the American Society of Civil Engineers. Editorial and General Offices at 33 West Thirty-ninth Street, New York 18, N. Y. Reprints from this publication may be made on condition that the full title of paper, name of author, page reference, and date of publication by the Society are given.

Copyright 1950 by the American Society of Civil Engineers
Printed in the United States of America

AMERICAN SOCIETY OF CIVIL ENGINEERS
Founded November 5, 1852
PAPERS

LONG-TERM STORAGE CAPACITY OF RESERVOIRS

By H. E. Hurst[1]

Synopsis

A solution of the problem of determining the reservoir storage required on a given stream, to guarantee a given draft, is presented in this paper. For example, if a long-time record of annual total discharges from the stream is available, the storage required to yield the average flow, each year, is obtained by computing the cumulative sums of the departures of the annual totals from the mean annual total discharge. The range from the maximum to the minimum of these cumulative totals is taken as the required storage.

The results of the investigations given here were applied in computing the storage required in the Great Lakes of the Nile River Basin, the length of time chosen as a basis of the estimates of storage being 100 years. These reservoirs, in combination with the other projects described, would enable the Nile to be developed for irrigation to the fullest possible extent. Finally, a relation has been derived between the storage capacity on the main stream and the amounts required on the tributaries to produce its equivalent.

It is thought that the general theory may have other applications than the design of reservoirs for the storage of water.

1. Introduction

A number of projects has been investigated which will enable Egypt and the Sudan to develop irrigation from the Nile River to its fullest extent.[2,3] The broad plan involves storage of water from good years for use in bad ones, and necessitates reservoirs of sufficient capacity to meet the shortages that might occur during a century. Much research had been done on the capacity

---

Note.—Written comments are invited for publication; the last discussion should be submitted by September 1, 1950.

[1] Scientific Consultant to the Ministry of Public Works, Egypt; formerly Director-Gen., Physical Dept., Cairo, Egypt.

[2] "The Nile Basin," Vol. VII, by H. E. Hurst, R. P. Black, and Y. M. Simaika, *Physical Department Paper No. 51*, Ministry of Public Works, Cairo Govt. Press, Cairo, Egypt, 1946.

[3] *Civil Engineering and Public Works Review*, September, 1948.

of such reservoirs and the results were used to determine the size of the reservoirs required in the Great Lakes of the Nile Basin for what was called "century storage."

Some preliminary results of investigations were released in 1938[4] which were amplified in 1946;[2] but in neither case was there an adequate statement of the theoretical investigation. The latter has been reserved for this paper.

Storage problems were treated first by the late Allen Hazen,[5] M. ASCE, in 1914, when he studied the discharge of thirteen American rivers, dealing with them mainly in a graphical manner, on probability paper. The longest record was one of 45 years, but longer sets of observations were made by combining records from several rivers. The objection to this procedure and the uncertainty in extending the normal Gaussian frequency distribution to the future were recognized by Mr. Hazen, but he thought it gave the best results then obtainable.

The work was extended by the late Charles E. Sudler,[6] M. ASCE, by similar graphical methods. Much of Mr. Sudler's work was based on artificial records in which information taken from a short record of a stream was extended to a period of 1,000 years by writing, say, 50 annual runoff values on cards and then shuffling and drawing a card from these 1,000 times. Again the actual data were necessarily confined to short periods since stream measurements in the United States had rarely been made for as long as 50 years. Thus, the results are mainly based on a few short-period records of natural river flow extended to long periods by random repetition. It is clear that the process can give very little information about what may happen over long periods and may be completely misleading since it excludes values higher or lower than those which have already occurred. This was recognized by some of those who contributed to the discussion on Mr. Sudler's paper. Moreover, the graphical methods employed do not lead to any concise and easily understandable presentation of results.

In the present investigation the methods of the theory of probability are applied, but the data are long-period records of natural events which are shown to have certain similarities, although they may be records of river discharge, rainfall, temperature, annual growth rings of trees, or annual deposits of clay in lakes. The average results can be expressed quite simply by two equations which apply to the storage of other substances as well as water. It is shown that, in regard to storage, records covering such short periods as 30 years or 40 years may be very misleading, but their usefulness can be extended by analyzing a large number of phenomena which have been recorded for periods as long as a century or more.

The problem of what storage is required on a stream to give a certain minimum discharge was first investigated in the case of Lake Albert (in the Belgian Congo and the Uganda Protectorate) to determine the size of an over-year storage reservoir which would be required in order to equalize the outflow over a number of years. It leads to a statistic R, which is the range from maximum to minimum of the curve obtained by plotting the cumulative totals of the departures from the mean of the annual discharges taken in order. Such a curve of cumulative totals is sometimes called a mass curve. Values of R have been computed for long series of observations of rainfall, temperature, and other natural phenomena, and these have been filed for reference in the Engineering Societies Library.[6a]

---

[4] "The Nile Basin," Vol. V, by H. E. Hurst and P. Phillips, Cairo Govt. Press, Cairo, Egypt, 1938, p. 86.

[5] "Storage to Be Provided in Impounding Reservoirs for Municipal Water Supply," by Allen Hazen, *Transactions*, ASCE, Vol. LXXVII, December, 1914, p. 1539.

[6] "Storage Required for the Regulation of Stream Flow," by Charles E. Sudler, *ibid.*, Vol. 91, 1927, p. 622.

Nomenclature.—The letter symbols introduced in this paper are defined where they first appear, in the text, or by illustration, and are arranged alphabetically in the Appendix, for convenience of reference. A few specialized terms used in the paper are defined as follows:

Century Storage.—The computed storage capacity of a reservoir, as required in the Great Lakes of the Nile Basin, necessary to meet the shortages that may occur in 100 years (Section 1).

Continued Sum.—Successive cumulative totals (Section 1).

Cubic Meter.—English units, 0.0008107 acre-ft.

Equatorial Lakes.—Lake Victoria (in Uganda, Tanganyika, and Kenya), Lake Albert (in the Belgian Congo and Uganda), and other bodies of impounded water in the southern Nile Basin, at the equator; also referred to as the "Great African Lakes," or simply the "Great Lakes" (Section 2).

Gaussian Frequency Distribution.—The normal frequency curve first advocated by the German mathematician, Karl Freidrich Gauss, who maintained that all skew distributions tended to become normal as the number of events was increased and that skewness was an evidence of inadequate data.

Meter.—English units, 3.28083 ft.

Milliard.—One thousand million things—that is, 10^9.

Varves.—Annual layers of silt as deposited in a lake or other body of still water. As individual layers differ in thickness and character, a succession of such layers forms a characteristic group which can be identified as of contemporaneous deposition in whatever deposit it may be found. It is thus possible, by combining different sections, to measure the time involved in the deposition of the entire group of sediments and to construct a time scale in a manner similar to that employed in the study of annual rings in trees (*Webster's New International Dictionary of the English Language*).

2. Statement of the Problem

Given a reservoir of large capacity, how can its outflow be regulated economically to meet the deficiencies of low years, and what is the capacity required to guarantee a certain minimum discharge in these years? These are the basic questions.

As a concrete example to illustrate the methods, consider the case of Lake Albert. If the past annual discharges from the lake are given, the size of reservoir required to maintain the maximum possible steady discharge during the period can be determined. This is the mean discharge for the period, and the capacity of the reservoir is obtained by adding, algebraically, the departures from the mean to form a series of accumulated totals. The difference between the highest and the lowest of these accumulated totals—that is, the range R—is (*a*) the maximum accumulated storage when there is never a deficit, (*b*) the maximum accumulated deficit when there is never any storage, or (*c*) their sum when there is both storage and deficit. This range R is the storage required to maintain the average discharge.

[6a] 29 W. 39th Street, New York 18, N. Y.

Increased losses due to storage are disregarded because, unless they are small, the site is not suitable for over-year storage. On a reservoir like the

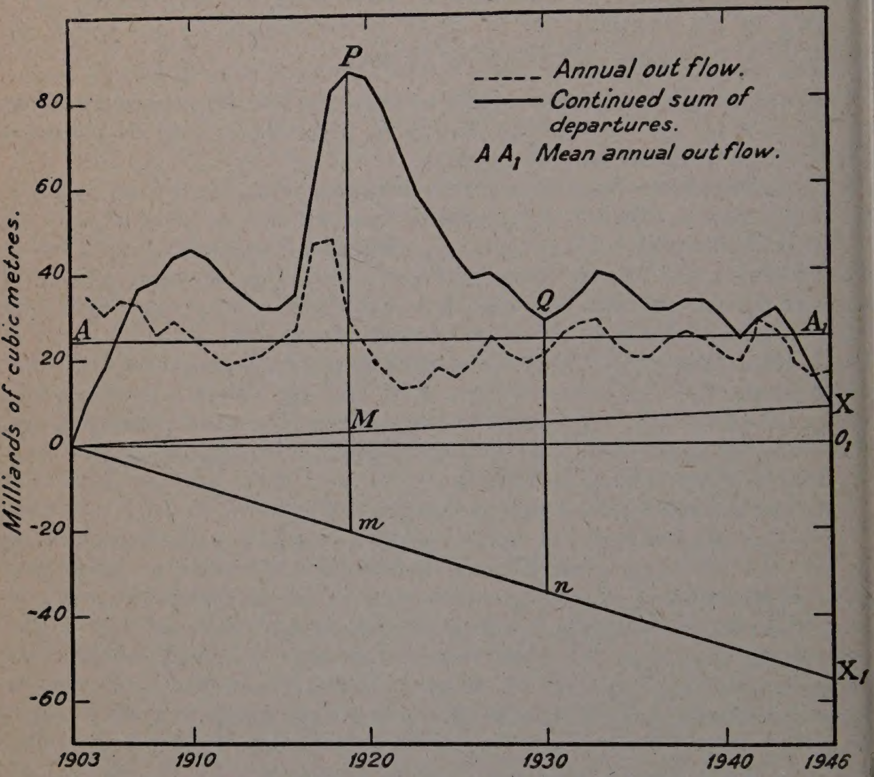

Fig. 1.—Annual Outflow from Lake Albert in the Belgian Congo and Uganda Protectorate (1904 to 1946) and the Curve of Continued Sums of Departures (One Milliard Cubic Meters = 810,700 Acre-Ft)

Aswan Reservoir in Egypt there is heavy evaporation uncompensated by any rainfall, and the surface area of the reservoir is several times that of the natural river. Therefore, there is an appreciable extra loss as a result of storage, so that, even if there were the capacity at Aswan, it would be unsuitable for over-year storage. In the case of the Equatorial Lakes their size makes it possible to store a great volume of water with only a small proportional increase of surface area. Moreover, evaporation and rainfall are nearly equal; so increased evaporation loss due to increased surface is compensated by increased gains from rainfall.

Fig. 1 shows the annual total discharge from Lake Albert based on recorded lake levels and gage-discharge curves. The curve of accumulated departures from a base near the mean is also shown. The actual values and the method of computation are shown by reference to Table 1.

TABLE 1.—DISCHARGE FROM LAKE ALBERT (BELGIAN CONGO—UGANDA PROTECTORATE) IN MILLIARD CUBIC METERS (1 MILLIARD CUBIC METERS = 810,700 ACRE-FT)

Year (1)	Q (2)	$+d$ (3)	$-d$ (4)	d^2 (5)	Σd_c (6)
1904	35	11		120	11
1905	31	7		50	18
1906	34	10		100	28
1907	33	9		80	37
1908	26	2		0	39
1909	29	5		20	44
1910	26	2		0	46
1911	22		2	0	44
1912	19		5	20	39
1913	20		4	20	35
1914	21		3	10	32
1915	24		0	0	32
1916	27	3		10	35
1917	47	23		530	58
1918	48	24		580	82
1919	29	5		20	87
1920	23		1	0	86
1921	17		7	50	79
1922	13		11	120	68
1923	14		10	100	58
1924	18		6	40	52
1925	16		8	60	44
1926	19		5	20	39
1927	25	1		0	40
1928	21		3	10	37
1929	19		5	20	32
1930	21		3	10	29
1931	26	2		0	31
1932	28	4		20	35
1933	29	5		20	40
1934	23		1	0	39
1935	20		4	20	35
1936	20		4	20	31
1937	24		0	0	31
1938	26	2		0	33
1939	24		0	0	33
1940	20		4	20	29
1941	19		5	20	24
1942	29	5		20	29
1943	26	2		0	31
1944	18		6	40	25
1945	15		9	80	16
1946	16		8	60	8
Σ	1,040	+122	−114	2,310	
N	43	(8)			
Mean	24.19	0.19			
Base	24				

The discharges, Q, in Col. 2, are usually given to one tenth of a milliard cubic meters but the last digit is not certain. For this computation the discharges are written to the nearest milliard and this reduces most of the departures, d, to a single digit, which is enough for purposes of computation. The omission of the last digit makes no significant difference to the mean or the standard deviation, σ, but lessens considerably the labor of computation. The computation is arranged to be self-checking as far as possible.

The mean of the series is first computed, and then a base is chosen near the mean. In this case the mean is 1,040/43, or 24.19, and 24 is selected as a base from which to compute departures, being the nearest whole number to the mean, and so keeping all the departures to one or two digits.

After writing the departures the positive and negative values of d (Cols. 3 and 4) are added separately. In this case the sum of the positive is 122 and of the negative, 114. The difference is 8 and $8/N$ should be the difference between the mean and the base. This latter difference gives a check on the correctness of the departures, and also on their continued sum Σd_c (Col. 6). If the departures were reckoned from the mean their final algebraic sum would be zero; otherwise, it is N (mean minus base), or 43 (24.19 − 24), and, of course, this is also the algebraic sum of the totals of the positive and negative departures, d, thus affording a check on the continued sums. With a long series of observations it is advisable to compute the totals of the columns, Σ, on each page to check the work piece by piece.

In computing the standard deviation, σ, the mean square deviation must be reduced by (mean minus base)^2, or by $(24.19 - 24)^2 = 0.19^2 = 0.04$, to give $\sigma^2 = \frac{2{,}310}{43} - 0.04$; and $\sigma = 7.33$. This is the only part of the work which is not self-checking and must be checked by repetition. The last digits of the square are omitted, which means in this case that the squares are given to the nearest ten which is near enough provided there are 50 or more observations. The effect of the approximation on σ is about 1% in this instance.

The departures are reckoned from a base which is not the mean, and, if the continued sums of these are plotted, a simple graphical method will yield the continued sum of departures from the mean or any other base. In the present case the mean is 24.2 and the base, 24. Every departure from the base is therefore 0.2 greater than the corresponding departure from the mean, so that the sum of the first n departures is $n \times 0.2$ greater than the sum of the first n departures from the mean. The sum of all the departures from the base is $N \times 0.2$. Thus, the continued sum of any number of departures from the mean can be found by joining the point representing the last continued sum to the origin and taking this line as a new axis. This is a useful construction for finding the continued sums from any base, when the sums referring to some base have been plotted.

In Fig. 1 the continued sums of departures from base 24 are plotted and the new axis OX giving continued sums of departures from the mean is drawn. The range, R, of the summation curve of departures from the mean is $\overline{\text{PM}} = 84$ milliards.

Examples of other uses of the graphic method of obtaining continued sums are given, subsequently, in Figs. 3 and 6.

In the case of Lake Albert the accumulated sum of departures rose to a maximum in 1919 (point P, Fig. 1) and then fell to zero by 1946 (point X), so that the next period would begin with no stored water. Had the operation been begun in 1920 there would have been a deficit to start with. This fact shows the necessity of making a reservoir some time before it will be required so that it can be partly or wholly filled before its storage is needed. It is clear from Fig. 1 that the storage required to maintain the mean discharge over the period is the range from maximum to minimum of the accumulated departures. It is also clear that, if the capacity of the reservoir is just enough to make up the maximum accumulated deficit, when this has been done, the reservoir will be empty and there will be no margin of stored water for the future. Moreover, the engineer is regulating the flow after the event, with full knowledge of what has happened. The problem of the required capacity is not a simple one and cannot be solved by adding departures from the mean, considering only Lake Albert, but must be dealt with on a wider basis. The discharge of the main Nile River at Aswan may be taken as a further illustration of the difficulties.

Table 2, for example, is based on three widely differing assumptions: First, that a computation was made in 1909, using 38 years of record; second, that a computation was made in 1946, using the succeeding 36 years of record; and, third, that a computation was made in 1946, using the total period of 74 years of record. The difference between the results of these computations shows how misleading the simple process can be, both as to the size of the reservoir needed and as to the benefit to be expected from it. Thus, it is clear that more extended investigation is required, and this can only be made by the examination of many sets of river discharges extending over long periods, or of other data with similar characteristics. At the same time some promise of a solution is possible by exploring theory.

TABLE 2.—COMPARISON OF VARYING RESULTS, COMPUTATIONS FOR THE MAIN NILE RIVER

Description	1871–1908	1909–1945	1871–1945
Average annual discharge, Q, milliards m^3............	103	83	93
Range of accumulated departures, R, milliards m^3......	201	83	476

The first part of this paper attempts to find a theoretical relation among the range, R, of the summation curve of departures from the arithmetic mean, the variability of the phenomenon expressed by its standard deviation, σ, and the number of years of observation, N, on the assumption that the observations are random events following the normal Gaussian frequency distribution. This course was adopted because many natural phenomena have frequency distributions which are approximately normal if the order of occurrence is disregarded, and it was thought that the examination of this case, which is a first approximation to those occurring in nature, would indicate the form of relation to be sought. Gauss' normal frequency curve (which is also the curve for the frequency distribution of errors of observation) gives the frequency of departures of a quantity from its arithmetic mean in relation to their magnitudes. If y is the number of departures lying between $x + \frac{1}{2}$ and $x - \frac{1}{2}$ when the total number of values of the quantity is N, the equation of the curve is

$$y = (N/\sigma \sqrt{2\,\pi})\, e^{-x^2/2\sigma^2} \ldots\ldots\ldots\ldots (1)$$

Incidentally, the assumption that the discharge of a river is a random event is the basis of the works of Messrs. Hazen[5] and Sudler;[6] under this same assumption a theoretical solution has been found.

3. THE FREQUENCY DISTRIBUTIONS OF RANDOM EVENTS

It is well known that the Gaussian normal frequency curve (Eq. 1) fits many distributions of events in which success or failure are equally likely and in which each event is independent of the next—for example, the frequency of accidental errors in a set of observations. It is also well known that an approximation to this curve is given by the terms of a binomial expansion. When a set of m coins is tossed, the probability of the occurrence of r heads and $(m - r)$-tails at any throw is ${}_mC_r\,(\frac{1}{2})^r\,(\frac{1}{2})^{m-r}$ in which ${}_mC_r$ is the number of combinations of m different things taken r at a time.

If the set is tossed N times (the number N being fairly large), the average frequency of occurrence of 0 heads and m tails, 1 head and $(m - 1)$-tails, etc., is given by the terms of the binomial expansion:

$$N\,(\tfrac{1}{2} + \tfrac{1}{2})^m = N\,(1 + {}_mC_1 + {}_mC_2 + \cdots + {}_mC_{m-1} + 1)\,(\tfrac{1}{2})^m \ldots\ldots (2)$$

As m increases, Eq. 2 approximates the normal curve, and if $m = 10$ the approximation is fairly good; in fact, for many purposes it is good enough.

Table 3 shows how the number of heads would be expected to be distributed on the average when a set of 10 coins is tossed 1,024 times—that is, the binomial distribution. The distribution derived from the normal curve, and one actually obtained by trial are also given.

TABLE 3.—COMPARISON OF FREQUENCIES

Distribution	Number of Heads, r											
	0	1	2	3	4	5	6	7	8	9	10	Total
Binomial	1	10	45	120	210	252	210	120	45	10	1	1,024
Normal Gaussian	2	10	43	116	212	258	212	116	43	10	2	
Actual experiment	0	16	53	122	209	240	194	138	46	5	1	

In the normal distribution the numerals represent the nearest integer. The relation between the binomial and the normal distribution has been treated by G. Udny Yule,[7] who defines the normal distribution as the limit of the binomial distribution when m is indefinitely increased. For theoretical purposes the computer can use whichever form is the more convenient.

4. CALCULATION OF THE RANGE OF CONTINUED SUMS OF DEPARTURES OR STORAGE FOR A NORMAL FREQUENCY DISTRIBUTION

Assume a number of consecutive values of an element whose variation from its mean is distributed normally, as in the binomial expansion already discussed, but where there is no correlation between successive values or groups of values, as in tossing coins. For example, consider the definite case of tossing a set of 2 m coins N times, calling a head a gain and a tail a loss. For present purposes each toss of the 2 m coins is recorded as the number of heads minus the number of tails, and the departures of this quantity from its mean are added to find the range R between the maximum and the minimum of this summation curve. The range thus found, from N tosses of a set of 2 m coins, may differ slightly from the range found by the summation of heads minus tails from 2 N m tosses of a single coin. This can be easily seen by tossing a set of 10 coins one after the other, and recording individual results. Then the sums of heads minus tails are the same at the end of every set of ten whether individual tosses are considered or not. It is as if the continued sum were plotted using single tosses, every tenth point being marked. The curve obtained by taking the tenth points would be a close approximation to the curve of individual points, but the latter would usually have a slightly greater range. For example, suppose that the maximum of the curve was produced by a set of 6 heads and 4 tails which would put the end-point of the set two units above the end-point of the preceding set. If the 6 heads occurred and then the 4 tails, there would be a point on the curve recording individual sums of heads minus tails which would be six above the end-point of the previous set—that is, the range would be increased by four. For the present this difference is ignored, but it will be considered later.

---

[7] "An Introduction to the Theory of Statistics," by G. Udny Yule, Charles Griffin and Co., London, 2d Ed., 1912, p. 317.

In the case selected for analysis $2\,m$ coins are tossed N times, heads and tails or gains and losses are equal at the end of the trial, and n is written for $N\,m$.

The number of orders or arrangements in which n gains and n losses can occur is

$$ {}_{2n}C_n = \frac{(2\,n)!}{(n\,!)^2} \qquad (3) $$

in which ${}_{2n}C_n$ is the number of combinations of $2\,n$ different things taken n at a time. Among these orders there are ${}_{2n}C_{n+h}$ instances in which, at some point in the process of tossing, losses exceed gains by h or more.[8]

There are three cases:

Case	Condition	No. of arrangements
(*a*)	Losses never exceed gains	$\dfrac{{}_{2n}C_n}{n+1}$
(*b*)	Gains never exceed losses	$\dfrac{{}_{2n}C_n}{n+1}$
(*c*)	Losses never exceed gains and gains never exceed losses	$(n-1)\,\dfrac{{}_{2n}C_n}{n+1}$

In case (*a*) the range is the maximum value of gains minus losses; in case (*b*) the range is the maximum value of losses minus gains; and in case (*c*) the range is the sum of the maximum of gains minus losses, and the maximum of losses minus gains.

Summing all values of the range to obtain a mean value, and writing G for number of gains and L for number of losses,

$$ \text{Mean range} = \frac{\text{maximum } (G-L) + \text{maximum } (L-G)}{{}_{2n}C_n} $$

$$ = \frac{2 \text{ maximum } (G-L)}{{}_{2n}C_n} \qquad (4) $$

To find the mean range see Table 4. There are ${}_{2n}C_{n+1}$ arrangements in which, at some point, gains exceed losses by one or more; and there are ${}_{2n}C_{n+2}$

TABLE 4.—COMPUTATIONS TO DETERMINE THE MEAN RANGE

Gains minus losses	No. of arrangements	Products
1	${}_{2n}C_{n+1} - {}_{2n}C_{n+2}$	$1\,({}_{2n}C_{n+1} - {}_{2n}C_{n+2})$
2	${}_{2n}C_{n+2} - {}_{2n}C_{n+3}$	$2\,({}_{2n}C_{n+2} - {}_{2n}C_{n+3})$
3	${}_{2n}C_{n+3} - {}_{2n}C_{n+4}$	$3\,({}_{2n}C_{n+3} - {}_{2n}C_{n+4})$
....	....	....
....	....	....
$n-1$	${}_{2n}C_{2n-1} - {}_{2n}C_{2n}$	$(n-1)\,({}_{2n}C_{2n-1} - {}_{2n}C_{2n})$
n	${}_{2n}C_{2n}$	$n\,({}_{2n}C_{2n})$
Sum	${}_{2n}C_{n+1}$	${}_{2n}C_{n+1} + {}_{2n}C_{n+2} + \cdots + {}_{2n}C_{2n}$

---

[8] "Choice and Chance," by W. A. Whitworth, Cambridge, England, 4th Ed., 1886, Chapter on Priority.

in which gains exceed losses by two or more. That is, there are ${}_{2n}C_{n+1} - {}_{2n}C_{n+2}$ arrangements in which gains exceed losses by one.

Referring to Table 4, $2^{2n} = 1 + {}_{2n}C_1 + {}_{2n}C_2 + \cdots + {}_{2n}C_{n-1} + {}_{2n}C_n + {}_{2n}C_{n+1} + \cdots + {}_{2n}C_{2n}$. Hence, the sum of the products is equal to $\frac{1}{2}(2^{2n} - {}_{2n}C_n)$. The arrangements in Table 4 include all except those in which gains never exceed losses, whose number is $\frac{{}_{2n}C_n}{n+1}$ (or ${}_{2n}C_n - {}_{2n}C_{n+1}$), and in which the maximum of gains minus losses is 0. Hence, the mean maximum value of gains minus losses is $\frac{2^{2n-1} - \frac{1}{2}\,{}_{2n}C_n}{{}_{2n}C_n}$ and this is also the mean maximum of losses minus gains. Therefore, the mean range is

$$R = \frac{2^{2n}}{{}_{2n}C_n} - 1 \qquad (5)$$

Since n is large, Eq. 5 can be simplified by the approximation for $n!$ of James Stirling which is as follows:

$$n! = \sqrt{2\,n\,\pi}\left(\frac{n}{e}\right)^n \qquad (6)$$

Hence,

$${}_{2n}C_n = \frac{(2\,n)!}{n!\,n!} = \frac{\sqrt{4\,n\,\pi}\,(2\,n/e)^{2n}}{[\sqrt{2\,n\,\pi}\,(n/e)^n]^2} = \frac{2^{2n}}{\sqrt{n\,\pi}} \qquad (7)$$

and

$$R = \sqrt{n\,\pi} - 1 = \sqrt{N\,m\,\pi} \qquad (8)$$

since the product $N\,m$ is large.

Thus the average range of the accumulated sums (number of heads minus number of tails), when $2\,m$ coins are tossed N times ($N\,m$ being large), increases as the square root of the number of tosses, exactly like the accumulated error on a line of leveling.

The standard deviation of the binomial distribution produced by tossing $2\,m$ coins is

$$\sigma_r = \sqrt{2\,m \times \tfrac{1}{2} \times \tfrac{1}{2}} = \sqrt{\frac{m}{2}} \qquad (9)$$

This is the standard deviation of the number of heads (r) or tails ($2\,m - r$). The standard deviation σ_d of the number of heads minus the number of tails is twice σ_r, or $\sqrt{2\,m}$. Substituting for $\sqrt{2\,m}$ in Eq. 8,

$$R = \sigma_d\sqrt{\tfrac{1}{2}\,N\,\pi} = 1.25\,\sigma_d\sqrt{N} \qquad (10)$$

In Eq. 10, R is the range of the continued sum of $2\,m\,N$ tosses of a single coin, and is a little larger than the range obtained from N tosses of a set of $2\,m$ coins. The average difference $\bar{d}$ between the two computations for range can be found as follows:

Toss a coin a large number of times and plot the continued sums of heads minus tails, marking a given number of them off in sets of $2\,m$. The set of

2 m that contains the maximum sum is selected for scrutiny. If there are r heads and t tails the end-point of the set is $r - t$ higher than the end-point of the previous set. The number of orders in which r heads and t tails can occur is ${}_{2m}C_r$.

The number of orders in which at some stage of the tossing the number of heads shall exceed the number of tails by h or more is ${}_{2m}C_{m+h}$. At the end of the set, heads exceed tails by $r - t = 2r - 2m$. The number of orders in which at some stage heads exceed tails by more than $2r - 2m$ is ${}_{2m}C_{2r-2m+1}$. In all these cases the maximum of the summation curve of individual tosses will exceed the maximum of the curve derived from sets of 2 m tosses. The mean value of the excess $\bar{d}$ can be obtained by a procedure similar to that already employed in finding the mean range of the summation curve. The result of this is

$$\bar{d} = \frac{1}{2^{2m}}\left[\frac{(2m)!}{(m-1)!\,(m+1)!} + 3\,\frac{(2m)!}{(m-2)!\,(m+2)!} + 5\,\frac{(2m)!}{(m-3)!\,(m+3)!} + \cdots + (2m-3)\,2m + (2m-1)\,1\right] \quad ..(11)$$

There is a similar effect on the minimum of the summation curve, so that the average range of the summation curve of individual tosses is greater than the range of the curve from sets of tosses by $2\bar{d}$. Some values of $2\bar{d}$ are given in Table 5. The correction $2\bar{d}$ to the range is independent of N and so has a less proportionate effect as N increases.

TABLE 5.—TYPICAL VALUES OF $2\bar{d}$

Magnitude	Values of m			
	4	5	6	10
$2\bar{d}$	1.5	1.7	2.0	2.7
Correction[a]	4.1	4.3	4.5	4.8

[a] Percentage correction to R when $N = 100$.

5. EXPERIMENTS WITH RANDOM EVENTS

Table 6(a) gives the results of an actual experiment in which sixpences were shaken in a box and thrown onto a table. They have been analyzed in some detail to give a practical example of the foregoing theory. Their frequency distribution has been given in Table 3, in which there were 1,024 throws of 10 coins ($m = 5$) which resulted in a distribution close to the theoretical.

The observations were recorded as they occurred and the continued sums were computed from 0 to 1,000 observations. In this case (because, at the finish, heads and tails are nearly equal) the actual values (heads minus tails) are taken as departures, and not the departures from the mean (Col. 1, Table 6(a)). At the finish there are 74 more heads than tails; so the mean value of $h - t$ per toss is 0.074. This is corrected graphically by the aforementioned method. Fig. 6, discussed subsequently, shows the result of one of the experiments.

In each case R (Cols. 3, Table 6) is corrected to what it would have been if the departure had been reckoned from the mean. The theoretical value of σ is 3.16 for a binomial distribution.

Table 6(*a*) shows that $R/(\sigma\sqrt{N})$ tends to be constant and the values from 1,000 trials (Cols. 4) agree closely with the theoretical values (Cols. 5). The trials illustrate a point not examined in the theoretical work, and that is the variability of σ, R, and $R/(\sigma\sqrt{N})$ from set to set of 100 throws.

The maximum variations, as percentage departures from the mean, are as follows: σ, 11%; R, 50%; and $R/(\sigma\sqrt{N})$, 39%. The standard deviation (or standard error) of σ for a set of 100 throws is about 7%.

Another device for obtaining a normal frequency distribution is what may be called a probability pack of cards (Table 6(*b*)). In this pack the cards are numbered +1, −1, +3, −3, +5, −5, +7, −7, +9, and −9, and the numbers of each kind are proportional to the corresponding ordinates of a normal frequency curve. They are: 13 one's; 10 three's; and 6, 3, and 1, respectively, of the others. The approximation of these numbers is fairly close. The cards are first well shuffled and then cut, and the number on the exposed card is recorded. The cards are reshuffled slightly and cut again, and so on. The numbers recorded may be taken as corresponding to observations of a quantity whose frequency distribution conforms to the normal Gaussian curve.

TABLE 6.—EXPERIMENTS WITH RANDOM EVENTS

DEFINITIONS OF COLUMN HEADINGS

Cols. 1. Primary Observations:
Table 6(*a*), Excess of heads minus tails = $h - t$.
Table 6(*b*), Sum of recorded numbers = Σ.
Table 6(*c*), Excess of even numbers over odd numbers = H.
Col. 2. Standard deviation of corresponding primary observations (Col. 1) = σ.
Col. 3. Range of accumulated deviations = R.
Col. 4. Computation of $\frac{R}{\sigma\sqrt{N}}$ from observed data.
Col. 5. Theoretical values of data in Cols. 4 corrected for the effect of tossing in sets.

No. of Trials	(*a*) Ten Sixpences Tossed 1,000 Times					(*b*) Probability Cards Cut 1,000 Times					(*c*) Serial Numbers of Bonds				
	$h - t$	σ	R	$R/(\sigma\sqrt{N})$		Σ	σ	R	$R/(\sigma\sqrt{N})$		H	σ	R	$R/(\sigma\sqrt{N})$	
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
1–100	26	3.08	39	1,28	1.20	−26	3.16	40	1.27	1.20	−10	3.42	39	1.14	1.20
101–200	24	3.10	25	0.81	....	+38	3.00	30	1.00	....	52	3.22	32	0.99	....
201–300	−26	3.36	44	1.31	....	− 2	3.99	35	0.88	....	− 4	2.30	24	1.04	....
301–400	−30	2.88	29	1.01	....	−44	4.06	36	0.89	....	−30	2.86	50	1.75	....
401–500	+24	3.59	40	1.11	....	+28	3.87	65	1.68	....	18	2.79	47	1.68	....
501–600	−32	3.06	31	1.01	....	− 2	3.58	48	1.34	....	6	3.13	56	1.79	....
601–700	+12	3.56	44.5	1.25	....	+10	3.57	33	0.92	....	8	2.97	46	1.55	....
701–800	+42	3.02	28.5	0.94	....	−22	3.11	44	1.42	....	30	2.78	44	1.58	....
801–900	+38	3.03	25.5	0.84	....	+ 4	3.82	78	2.04	....	−48	3.02	49	1.62	....
901–1,000	− 4	3.50	54	1.54	....	−14	4.25	48	1.13	....	8	3.44	40	1.16	....
Mean	....	3.22	36.1	1.11	1.20	....	3.64	45.7	1.26	1.20	....	2.99	42.7	1.43	1.20
1–500	+18	3.20	87	1.21	1.23	− 6	3.62	81	1.00	1.23	26	2.95	84	1.27	1.23
501–1,000	+56	3.23	94	1.30	1.23	−24	3.67	111	1.35	1.23	4	3.09	87	1.26	1.23
1–1,000	+74	3.22	121	1.19	1.23	−30	3.64	116	1.01	1.23	30	3.02	88	0.92	1.23
Mean	....	....	....	1.19	1.22	....	....	....	1.15	1.22	....	....	....	1.20	1.22

This is a quicker process than tossing and counting coins. To toss 10 coins 100 times required about 35 min, whereas shuffling and cutting for 100 cards required 20 min. Table 6(*b*) records the results of 1 trial of cutting 1,000 times. As in Col. 3, Table 6(*a*), in each case R is corrected to what it would have been if the departures had been reckoned from the mean; and, again, $R/(\sigma\sqrt{N})$ in Cols. 4 approximates the value obtained theoretically for the case when N is large (Col. 5).

A third set of chance events was examined (Table 6(*c*)). From time to time newspapers announce the numbers of bonds that have been drawn for redemption. A collection of unconnected random events was made by dividing the numbers of the bonds into groups of ten, as follows:

	(*a*)	(*b*)	(*c*)		(*d*)	(*e*)	(*f*)
3	5	7	0	3	6	0	3
3	8	1	0	3	8	7	6
3	9	8	9	3	9	9	1
4	1	0	5	4	1	1	7
4	2	4	5	4	2	5	9
4	4	0	4	4	4	1	6
4	5	3	4	4	5	3	6
4	6	6	8	4	6	7	1
4	7	7	3	4	8	0	0
4	9	2	2	4	9	2	3
	–	–	–		–	–	–

Number of Even Digits Minus Number of Odd Digits

−2	+0	+2	2	−4	−2

The left-hand column of a group was ignored as only a few different digits are likely to occur in it. The number of even digits less the number of odd digits in columns (*a*) (*b*) (*c*), (*d*) (*e*) (*f*), etc., are recorded. This difference yields a sequence of numbers exactly analogous to the sequence previously obtained by tossing coins, and the labor of producing it is much less.

By this method a sequence of 1,000 random numbers composed of the even numbers from 0 to ±10 was made and this was used as before. Table 6(*c*) gives the results of this trial.

In each case R (Col. 3) has been corrected to what it would have been if the departures had been reckoned from the mean. The theoretical value of $R/(\sigma\sqrt{N})$ (Col. 5) has been corrected for the effect of tossing in sets.

Although the mean value of $R/(\sigma\sqrt{N})$ found from the three sets of random events in Table 6 is very close to the value found by mathematical analysis (Cols. 5 compared to Cols. 4) yet individual values for sets of 100 observations vary considerably. The extreme ranges in the three cases are as follows:

Table	$\frac{R}{\sigma\sqrt{N}}$
6(*a*)	0.81 to 1.31
6(*b*)	0.88 to 2.04
6(*c*)	0.99 to 1.79

The mean value of $R/(\sigma\sqrt{N})$ from the thirty sets of 100 observations is 1.27 and the standard deviation is 0.32. The relation between R, σ, and N for random events having been established theoretically and having been found to agree with practice, it is now necessary to find in what respects natural phenomena are similar to random events and in what respects they differ.

6. CHARACTERISTICS OF NATURAL PHENOMENA

The phenomena that chiefly concern the subject of this paper are river discharges. Unfortunately records of these rarely cover periods of more than 50 years, although levels (which are usually an indication of the discharge) have been recorded over longer periods. The discharge of a river is closely related to the rainfall on its basin, and rainfall records for many places cover more than 100 years. The investigation began with these phenomena.

The main similarity of river discharges, levels and rainfall, and random events is that their frequency distributions conform approximately to the normal curve, although they are usually slightly skew. As a rule the observations are too few to give very regular distributions, but the flood levels (nearly 1,000 of them), of the Nile River at Cairo, Egypt, recorded on the Roda gage, give a very regular distribution which is well fitted by a normal curve. This distribution is shown in Fig. 2.

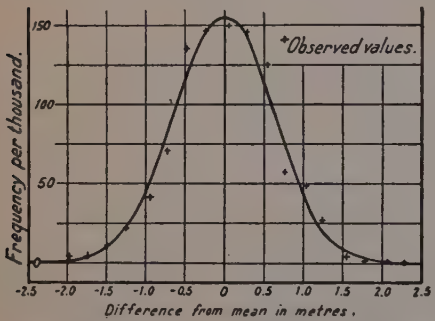

FIG. 2.—MAXIMUM ANNUAL GAGE READING AT RODA GAGE, CAIRO, EGYPT, FOR 1,080 YEARS BETWEEN 641 A.D. AND 1946 A.D.

Although in random events groups of high or low values do occur, their tendency to occur in natural events is greater. This is the main difference between natural and random events, an example of which is the discharge of the Nile River at Aswan. The long series of records of flood levels at Cairo shows the same phenomenon.[9] There is no obvious periodicity, but there are long stretches when the floods are generally high, and others when they are generally low. These stretches occur without any regularity either in their time of occurrence or duration.

In the investigations of Messrs. Hazen and Sudler the frequency characteristics of river discharges are assumed to be like those of random events, and this is the common assumption in estimating probability of high or low floods. It will be shown that this assumption must be used with caution and, in cases in which storage over long periods is concerned, it is only an approximation.

7. THE RANGE (R) OF THE CONTINUED SUMS OF DEPARTURES FOR NATURAL PHENOMENA

The investigation began experimentally, before any attempts to find a mathematical theory were made, by finding R for any available long series of river discharges. Such series were scarce and so the work was extended to rainfall data, which are more plentiful. When it was found that rainfall data gave results similar to the river data the work was extended to other long series of natural events.

---

[9] "Flood-Stage Records of the River Nile," by C. S. Jarvis, *Transactions*, ASCE, Vol. 101, 1936, p. 1012.

For example, Fig. 3 shows a summation curve of departures for the maximum levels of a lake on the Dalalven River in Sweden. There are 176 years of observations, divided into periods of 44 years, 88 years, and 132 years; and the ranges, R, of the summation curves for the departures from the mean in each of these periods are found.

The sums of departures from the base OX are plotted. For the sum of departures from the mean for the entire period the curve is referred to the axis OD. Similarly, the summation for the first 44 years is referred to line OA, for the second 44 years to line AB, for the first 88 years to line OB, and so on. The range, R, of the last period BD is $\overline{QN} + \overline{PM}$.

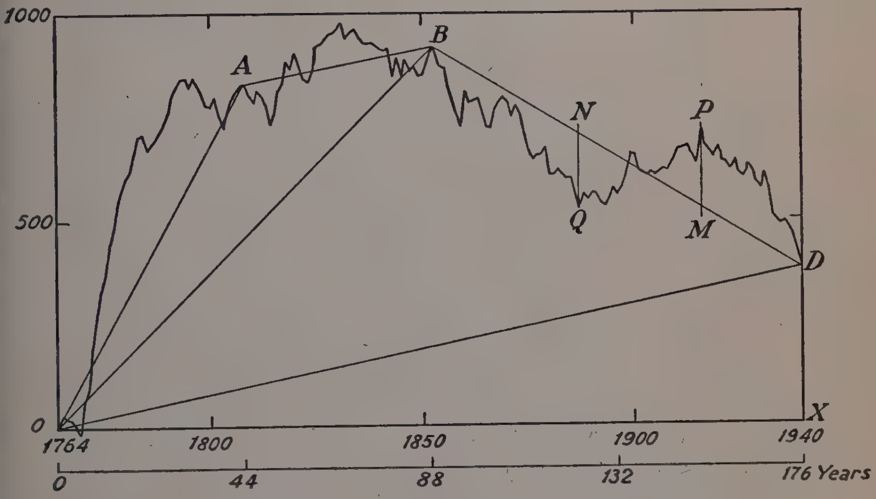

Fig. 3.—Summation Curve of Departures, Maximum Levels of a Small Lake on the Dalalven River in Sweden

8. Variation of R with N and σ

The theoretical investigation shows that R is proportional to σ in the case of random events, and this must be approximately the case for any events that have normal or nearly normal frequency distributions. The device of expressing deviations from the mean in terms of σ is a common one, derived from the expression for the normal curve; and, since R is a function of these deviations, R/σ has been calculated. Complete tabular records of the computation including mean values of N and K $\left(= \dfrac{\log (R/\sigma)}{\log (N/2)}\right)$ are in the complete original manuscript, cataloged for reference in the Engineering Societies Library.[6a] Seventy-five different phenomena were used, and 690 values of R were computed for different values of N.

A summary of the results is given in Table 7 in which the phenomena are divided into groups according to their nature and the value of N, and log (R/σ) is plotted against log N in Fig. 4.

In general, R has not been computed for periods of less than 30 years because, for very short periods, it has little practical value although it may have some formal value. For $N = 2$, R/σ has the exact value, unity. The longest period used is 2,000 years and this is for clay layers in Lake Saki in the Crimea. Fig. 4, in which the means of small groups of observations of the same kind and length of period have been plotted, shows that log (R/σ) is a linear func-

TABLE 7.—RELATION OF R/σ AND N FOR GROUPS OF PHENOMENA

Phenomena	No. of cases	N years	$\frac{R}{\sigma}$	Log N	Log $\frac{R}{\sigma}$	K
(a) GROUP OF 99 CASES						
River levels, discharges, and runoff	8	35	7.5	1.54	0.85	0.68
	8	45	8.9	1.65	0.94	0.70
	8	62	13.1	1.79	1.08	0.72
	9	108	16.4	2.02	1.19	0.69
Roda nilometer	12	105	19.6	2.02	1.27	0.75
	10	208	36.5	2.32	1.54	0.77
	9	309	53.9	2.49	1.72	0.79
	8	420	60.3	2.56	1.77	0.78
	7	511	67.7	2.71	1.82	0.75
	6	613	81.3	2.79	1.89	0.77
	5	716	104	2.85	2.01	0.79
	4	820	122	2.91	2.08	0.80
	3	927	129	2.96	2.11	0.79
	2	1,040	130	3.02	2.12	0.78
Mean of 99 cases	....	....	....	2.28	1.48	0.75
(b) GROUP OF 168 CASES						
Rainfall stations with one value of R^a	7	71	14.0	1.82	1.12	0.74
Two groups of values of R^b	20	48	8.7	1.70	0.92	0.66
	12	98	13.7	1.94	1.11	0.68
Rainfall stations with three groups of values of R	66	38	8.2	1.57	0.91	0.72
	38	78	14.4	1.87	1.13	0.72
	25	121	22.2	2.08	1.31	0.74
Mean of 168 cases	....	....	....	1.79	1.08	0.70
(c) GROUP OF 109 CASES						
Temperature and pressure stations with 3 values of R	27	37	7.4	1.55	0.85	0.68
	18	73	12.0	1.86	1.06	0.68
	10	110	18.0	2.04	1.24	0.71
Temperature stations with 4 values of R	24	44	8.8	1.63	0.93	0.70
	12	88	15.1	1.94	1.17	0.71
	12	132	21.0	2.12	1.32	0.72
	6	175	27.1	2.24	1.43	0.74
Mean of 109 cases	....	....	....	1.81	1.05	0.70
(d) GROUP OF 85 CASES						
Annual growth of tree rings	41	50	13.6	1.69	1.11	0.80
	21	100	26.6	2.00	1.43	0.84
	14	200	45.0	2.30	1.64	0.82
	4	300	79.9	2.48	1.90	0.87
	4	462	76.6	2.66	1.87	0.79
	1	900	187.0	2.95	2.27	0.86
Mean of 85 cases	....	....	....	1.97	1.36	0.80

TABLE 7.—(*Continued*)

Phenomena	No. of cases	N years	$\frac{R}{\sigma}$	Log N	Log $\frac{R}{\sigma}$	K
(*e*) GROUP OF 90 CASES						
Thickness of annual layers of mud; Tamiskaming, Ont., Canada, and Moen, in the Sogne District, Norway..............	44	50	10.9	1.70	1.02	0.73
	22	100	21.3	2.00	1.31	0.77
	11	200	42.7	2.30	1.58	0.79
	4	300	82.5	2.48	1.90	0.87
	3	400	126	2.60	2.09	0.91
	4	550	115	2.74	2.00	0.82
	2	1,100	181	3.04	2.19	0.80
Mean of 90 cases....................	....	....	....	1.98	1.30	0.77
(*f*) GROUP OF 114 CASES						
Thickness of annual layers of mud, Lake Saki in the Crimea..................	40	50	9.7	1.70	0.98	0.70
	40	100	15.3	2.00	1.17	0.69
	20	200	25.0	2.30	1.39	0.70
	8	500	47.9	2.70	1.66	0.69
	4	1,000	84.0	3.00	1.91	0.71
	2	2,000	179.0	3.30	2.24	0.75
Mean of 114 cases....................	....	....	....	2.06	1.22	0.69
(*g*) GROUP OF 25 CASES						
Sunspot numbers and wheat prices.......	12	64	12.4	1.77	1.06	0.72
	6	124	22.1	2.09	1.34	0.75
	7	237	16.9	2.36	1.22	0.60
Mean of 25 cases....................	....	....	....	2.01	1.17	0.69
(*h*) GROUP OF 259 CASES						
Weight mean of 259 cases[c]..............	....	97	17.8	....	....	....

[a] Rainfall stations with one value of R; includes temperature at one station. [b] Rainfall stations with two groups of values of R; includes temperature and one pressure. [c] N ranges from 81 to 120.

tion of log N and can be expressed as

$$\log \frac{R}{\sigma} = K \log \frac{N}{2} \qquad (12)$$

In each main group of observations a line drawn through the point $\log (R/\sigma) = 0$, $\log N = 0.30$ ($N = 2$), and the center of gravity of all the observations is a good fit. The slopes of these lines (K) vary from 0.69 to 0.80.

Whether there is any theoretical significance in the fact that K is approximately $\frac{3}{4}$ is not known. The value of K has been calculated for each of the 690 individual values of R/σ and the frequency curve shown in Fig. 5 has been drawn. The data for this curve are as follows:

Description	K
Number of values..........................	690
Mean value..........................	0.729
Standard deviation..........................	0.092
Range of individual values..................	0.46 to 0.96

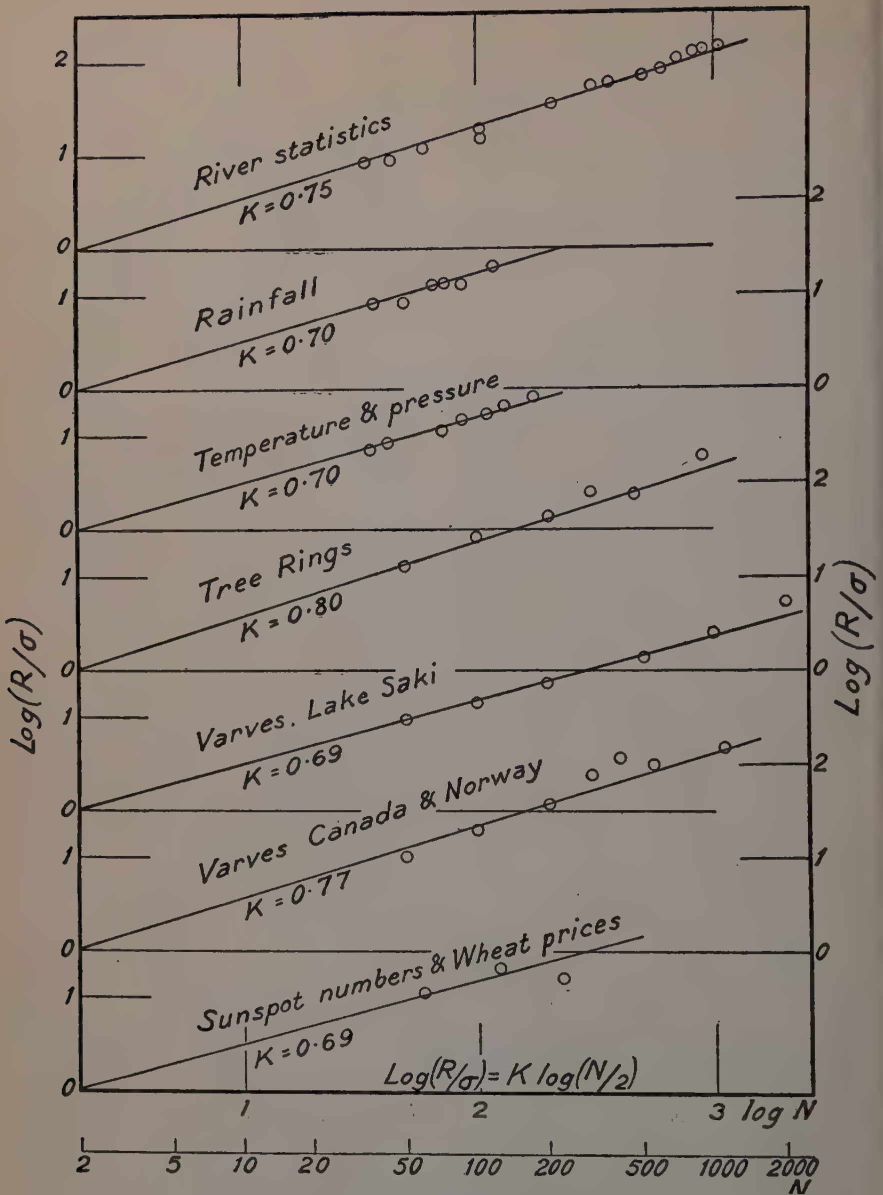

FIG. 4.—RELATION BETWEEN RANGE OF SUMMATION CURVE R, THE STANDARD DEVIATION σ, AND THE YEARS OF RECORD N

Fig. 5 shows that there is a slight skewness in the distribution of K, the mode tending to be slightly greater than the mean. However, it will be seen that the normal curve which has been fitted is a close approximation to the observed distribution. Referring to Table 7, a summary of K-values can be compiled as in Table 8.

Table 8 and Fig. 5 show a remarkable result—that the values of K are nearly the same regardless of the class of phenomenon from which they are determined. It is necessary to examine them a little further to see if the differences are significant. It will be noticed that the standard deviation of K from all the values is a little larger than that for most of the values for the groups. When the standard deviation of the means of groups is calculated it is found to be about 4.25 times the standard deviation of the mean of a similar group of random values of K. This suggests that K does vary slightly with different phenomena. The largest mean value of K is 0.81 from tree rings, and this is derived from four sets of trees, each set being from one of four different places. The means of sets are 0.84, 0.80, 0.81, and 0.81, so that it would appear that the departure of the trees from the mean of all the phenomena is significant. Similarly, the difference in K between the varves of Lake Saki (the Crimea) and those of Tamiskaming (Ontario, Canada) and Moen (Norway) is greater than would be expected on a chance distribution. Tamiskaming and Moen are significantly different. The rainfall observations have been examined in detail, with the result that the means for fifteen stations with some 7 values of K to each station do not differ greatly from a random distribution.

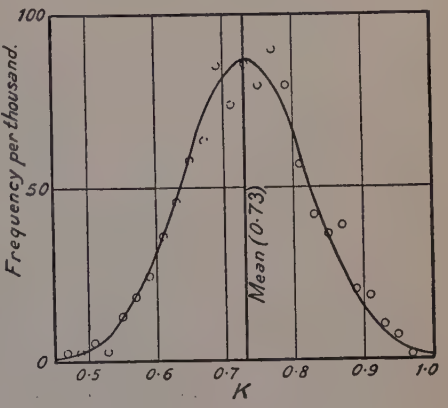

FIG. 5.—FREQUENCY OF INDEX K

TABLE 8.—SUMMARY OF K-VALUES

Phenomena	No.	Mean	Standard deviation
River levels, discharges, etc.	99	0.75	0.077
Rainfall	168	0.70	0.069
Temperature and pressure	115	0.70	0.085
Annual growth of tree rings	85	0.81	0.078
Varves (Lake Saki in the Crimea)	114	0.69	0.064
Varves (Tamiskaming, Ont., Canada, and Moen, Norway)	90	0.77	0.094
Sunspot numbers and wheat prices (combined as miscellaneous phenomena)	25	0.69	0.086
Means and totals	690	(0.729)	(0.092)

The general conclusion is that, if a long series of several hundred years is available for a phenomenon, it is probably best to take the mean value of K from these as an indication for the future. However, if only a short period of from 100 years to 200 years or less is available, the best plan is to take the mean value for K from all available material. The mean adopted for this purpose is 0.72 instead of 0.73, in order to give a little more weight to the more precise measurements relating to rivers, rainfall, and temperature. The mean value adopted, therefore, is $K = 0.72 \pm 0.006$. In using this result it is to be noted that it is the mean value of a quantity that has ranged from 0.46 to 0.96, and, therefore, some margin of safety must be allowed.

The probable error of a single value of R/σ can be derived by differentiating Eq. 12 with respect to K and using the data in Table 8, which yield a probable error of 24%. Confirmation of this value is obtained by taking the values of R/σ for rainfall, river discharge, temperature, and pressure for values of N between 80 years and 120 years, and correcting them to a period of 100 years. The probable error calculated from these corrected values is 23%. Fig. 10, introduced subsequently, shows how R/σ can be reduced by 23%, by reducing the draft by 0.04 σ.

In Fig. 5. the mean representing all the lines is represented by

$$\frac{R}{\sigma} = \left(\frac{N}{2}\right)^{0.72} = 0.61\,N^{0.72} \qquad (13)$$

which means that R/σ increases more rapidly with N in the case of natural phenomena than in the case of chance events for which the power of $N = \frac{1}{2}$. The value of R/σ for $N = 100$ from Eq. 13 is 16.8. In the computations previously cited[2] the equation $R/\sigma = 1.65\sqrt{N}$ was used, which gives $R/\sigma = 16.5$ for $N = 100$.

If a phenomenon is exactly periodic, with constant amplitude, R will have a definite maximum which will be reached at the end of each complete period. In this case R will not vary in the manner already found for natural phenomena. The phenomena considered in this paper, however, are not of this type. In the case of sunspot numbers (which are the nearest approach) the periodic part does not have a constant amplitude, and has irregular variations superposed. These irregularities cause R to increase as N increases and the mean value of K is 0.75, with no indication that the phenomenon is different from the many others considered.

The very striking result from the investigation is the small variation of the index K over such a wide range of phenomena, and hence the possibility of its utility in many fields not mentioned in this paper.

9. THE VARIATION OF THE MEAN AND THE STANDARD DEVIATIONS DERIVED FROM SHORT-TERM RECORDS

When records of a natural phenomenon extend over long periods there are considerable variations of both means and standard deviations from one period to another. The primary practical interest of this paper is in water supply and, hence, in rainfall. Unfortunately 200 years is about the longest series of observations obtainable for rainfall, and few stations reach 150 years. Some of these records are given in Table 9. They are divided into periods of 40 years or 50 years to illustrate the recorded variations. This division is chosen because quite frequently water supply or hydroelectric projects have to be based on such short records since nothing more is available. The reservoirs in the Great African Lakes are a case in point.

The average range of variation of the mean values of the rainfall in Table 9 is about 14%, whereas the standard deviations vary on the average by 23%.

This is more than would be expected from the ordinary theory for random distributions. The same is true in the case of temperature.

The uniformity in the values of R shown in the preceding section encourages the examination of some of the long-period records, a summary of which is compiled in Table 10. In Table 10(a) the variability of means for 50 years and 100 years is compared with what would be expected if the frequency were random, and not affected by any tendency for high or low values to be grouped together.

TABLE 9.—VARIATION OF MEANS AND STANDARD DEVIATIONS IN THE CASE OF RAINFALL

Station	Duration of records (years)	Length of period (years)	No. of periods	Mean		Standard Deviation	
				Maximum	Minimum	Maximum	Minimum
Philadelphia, Pa.	126	42	3	42.7	41.4	7.0	5.6
Boston, Mass.	128	43	3	44.2	38.2	7.4	5.0
Madras, India	133	44	3	51.3	48.4	15.6	14.3
Rome, Italy	151	50	3	886	777	17.1	15.7
Stockholm, Sweden	161	40	4	57.8	38.3	13.5	7.8
Padua, Italy	171	43	4	940	793	179	147
Milan, Italy	171	49	6	104	95	21.2	15.1
Zwanenbourg, the Netherlands	211	53	4	77	69	13.1	10.8

The computed value of the standard deviation (σ) of a mean for 50 years (Col. 4) is obtained by taking the standard deviation obtained from all the individual observations for the period given in Col. 3 and dividing it by $\sqrt{50}$. Cols. 5, Table 10, are the ratios Cols. 3/Cols. 4. It appears from Table 10(a) that the means are much more variable than they would be in the case of a random distribution, the standard deviation of a 50-year mean being 2.5 times

TABLE 10.—VARIATIONS OF MEANS AND STANDARD DEVIATIONS

Phenomenon (1)	No. of years, N (2)	(a) Standard Deviations of Means						(b) Standard Deviations of Standard Deviations					
		50-Year Means			100-Year Means			50-Year Periods			100-Year Periods		
		Actual (3)	Computed (4)	Ratio (5)	Actual (3)	Computed (4)	Ratio (5)	Actual (3)	Computed (4)	Ratio (5)	Actual (3)	Computed (4)	Ratio (5)
Varves:													
Lake Saki, Crimea	1,000	2.0	0.9	2.2	1.8	0.6	2.8	2.6	0.6	4.2	3.1	0.5	6.5
	1,000	1.5	0.9	1.7	1.0	0.6	1.5						
	1,000	2.2	1.3	1.7	1.6	0.9	1.7						
	950	2.2	1.2	1.8	2.0	0.9	2.3						
Moen, Norway	1,000	4.1	1.9	2.1	3.5	1.4	2.5	5.6	1.2	4.8	4.5	0.9	5.0
Tamiskaming, Ont., Canada	1,200	4.2	0.8	5.5	3.5	0.6	6.2	1.11	0.36	3.1	1.6	0.3	5.4
Ring Thickness:													
Sequoia trees	900	0.12	0.04	2.8	0.10	0.02	5.6	0.034	0.014	2.4	0.032	0.011	2.9
Flagstaff pines	500	0.09	0.04	2.1	....	....	....	0.040	0.029	1.4	....	....	....
Roda gage (free of trend)	1,050	0.24	0.10	2.4	0.21	0.07	3.0	....	....	....	0.145	0.042	3.4
Means	....	....	....	(2.5)	....	....	(3.2)	....	....	(3.2)	....	....	(4.6)

as great as it would be for a random distribution. Similarly, the standard deviation of a 100-year mean is 3.2 times as great as in a random distribution. This implies that, over a long period, the maximum value is likely to be higher and the minimum lower than would be predicted from the application of the ordinary theory of probability to the records from a short period.

Table 10(*b*) shows the variability of the standard deviation for periods of 50 years and 100 years in those cases where the records cover a long time. If a distribution is normal and randomly distributed as in chance events, the standard deviation of the standard deviation is $\sigma/\sqrt{2N}$. This is the value called "computed" (Cols. 4). The actual value (Cols. 3) is found in the usual way from the standard deviations for all the 50-year periods, etc., and Cols. 5 contain the ratios, Cols. 3/Cols. 4. Again it appears that the standard deviation for these natural phenomena is more variable than would be the case in a random distribution.

Summary.—Although many natural phenomena have a nearly normal frequency distribution, this is only the case when their order of occurrence is not considered. Consequently, their representation by a random distribution in the usual way is only a first approximation. The tendency to occur in groups makes both the mean and the standard deviation computed from short periods of years more variable than in the case of random distributions. Thus the odds against the occurrence of some particular extreme value (say, a high flood) are likely to be less than those calculated from a short period of records on normal probability theory.

It was thought that it might be possible to derive an expression for the effect of grouping, involving the correlation coefficient between successive values of the variable. Experiments with the natural number $1, 2, \cdots, N$ and $-1, -2, \cdots, -N$, however, showed that this was not the case. The greatest value of R is obtained by placing the positive numbers in one group and the negative in another and is independent of the order in the groups. The correlation between successive numbers depends on the order within the groups. In the foregoing example, the coefficient of correlation may vary from 1 to $\frac{1}{2}$ while R remains constant at $N(N+1)$. The numbers can also be arranged to give small values of R with correlation coefficients varying from $\frac{1}{2}$ to -1.

10. STORAGE TO GUARANTEE A DRAFT LESS THAN THE MEAN DISCHARGE

If the discharge from a reservoir over a period is kept steady at the mean, the range of the curve of progressive sums of departures from the mean is taken as the storage required to keep this constant discharge. The final sum of the departures is zero. If departures are taken from a base less than the mean, this corresponds to the case in which the reservoir discharge is not allowed to fall below this base value.

The storage required in this case is the greatest accumulated deficit. This is illustrated in Fig. 1, of Lake Albert discharges, in which the summation curve, or curve of accumulated departures from base 24, is plotted, axis OO_1, Fig. 1, and the summation curve for departures from the mean is referred to axis OX.

If departures are computed from a new base 0.2σ below the mean, as explained previously, the summation curve must be referred to the inclined axis OX_1, where the ordinate XX_1 is $43 \times 0.2\sigma$ or 63 milliards. It is obvious that, as the inclination of the axis OX_1 changes, different points on the summation curve may become maxima, or minima, and there is no simple connection between the storages required to guarantee the mean and to guarantee some lesser discharge. This is clearly shown in Fig. 6, in which the result of tossing 10 coins 100 times is shown in the form of a curve of the continued sums of the number of heads minus the number of tails.

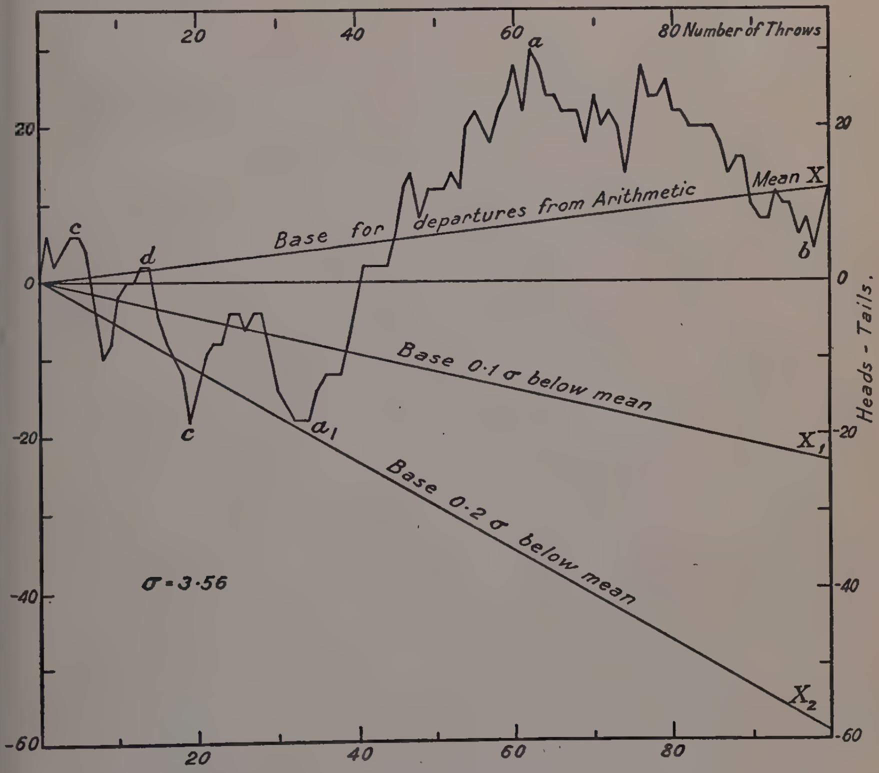

Fig. 6.—Summation Curve of Heads Minus Tails, in Tossing 10 Coins 100 Times

At the end of the 100 tosses, heads exceeded tails by twelve, so that the departures of heads minus tails are reckoned from a base 0.12 below the mean. This is corrected in Fig. 6 by drawing the axis OX which is the base for the summation curve of departures of heads minus tails from their mean. Similarly, the axes OX_1 and OX_2 can be drawn for the summation curves of departures from the bases 0.1σ and 0.2σ below the mean. The two points aa_1 are the maximum and minimum of the summation curve of heads minus tails and also of the summation curve of departures of heads minus tails from its mean. The range of the first curve is greater than the range of the second.

The maximum accumulated deficit with the mean as base is given by the points ab, but when a base 0.1 σ is selected below the mean the maximum accumulated deficit is given by cc_1; for a base 0.2 σ below the mean, it is dc_1. The storage required to maintain a draft below the mean can be computed arithmetically; but it is extremely laborious and by far the quicker and more reliable method is the graphic one.

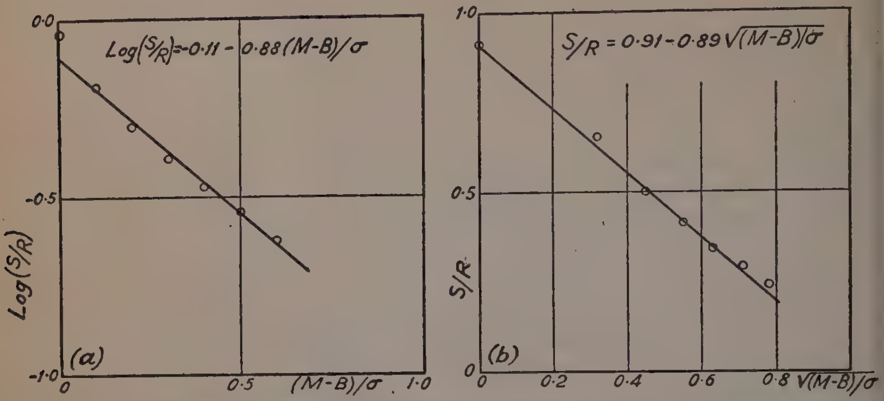

FIG. 7.—ANALYSIS OF TOSSED COINS, CARDS, AND DRAWING OF BONDS TO DETERMINE A RELATION BETWEEN DRAFT B AND MAXIMUM DEFICIT S (SEE TABLE 11)

TABLE 11.—VALUES OF S/R FOR GIVEN RATIOS $(M - B)/\sigma$ (MEANS OF 10 EXPERIMENTS OF 100 TRIALS EACH)

Media	R	VALUES OF $(M - B)/\sigma$					
		0.1	0.2	0.3	0.4	0.5	0.6
Coins	36	0.62	0.47	0.40	0.34	0.31	0.26
Cards	46	0.71	0.56	0.45	0.36	0.29	0.23
Bonds	43	0.62	0.48	0.39	0.32	0.27	0.23
Means	42	0.65	0.50	0.41	0.34	0.29	0.24
Logs		−0.19	−0.30	−0.39	−0.47	−0.54	−0.62

TABLE 12.—RELATION BETWEEN DRAFT B A

Magnitude	VALUES									
	0.06	0.09	0.1	0.20	0.23	0.26	0.29	0.30	0.37	0.4
(a) RIVERS, RAINFALL, AND TEMPERATU										
S/R	0.76	0.77	....	0.56	0.49	0.47	0.46	0.42	0.32	...
Log (S/R)	−0.12	−0.11	....	−0.26	−0.32	−0.34	−0.37	−0,40	−0.51	...
(b) TREE RINGS AND VARVES, RODA G										
S/R	....	....	0.64	0.48	....	....	....	0.35	....	0.
Log (S/R)	....	....	−0.19	−0.32	....	....	....	−0.46	....	−0.

11. THE RELATION BETWEEN DRAFT AND STORAGE FOR CHANCE EVENTS

Let the storage required to maintain a draft B less than the mean (M) be S. Values of S (the maximum deficit) have been computed from the experiments with coins, probability cards, and drawing of bonds, as described. They have been calculated for every set of 100 experiments and for values of $M - B$ of $0.1\ \sigma$, $0.2\ \sigma$, etc. The ratio S/R has been tabulated and mean values are given in Table 11.

The results are also given in Fig. 7(*a*) in which $\log S/R$ is plotted against $(M - B)/\sigma$, the relation being

$$\log \frac{S}{R} = -0.11 - 0.88\,\frac{M - B}{\sigma} \qquad (14a)$$

A little better fit is obtained from the equation—

$$\frac{S}{R} = 0.91 - 0.89\sqrt{\frac{M - B}{\sigma}} \qquad (14b)$$

which is shown in Fig. 7(*b*). Neither formula should be used outside the range of values from which it was computed.

Considerable research was devoted to finding a theoretical relation between S and R, but without success. An extension of the method of Section 5 will enable the mean range of the summation curve of departures from a base below the mean to be found, but it does not result in a simple form. This mean range is not what is wanted, however; it is the maximum accumulated deficit in the entire period, or any part of it that is wanted. The latter is a complicated function of gains minus losses, for which no algebraic expression has been found.

An interesting point in the case of coins, cards, etc., is that, if the mean range is called R_1 when departures are measured from a base less than the mean, then the ratio R/R_1 is approximately equal to S/R.

ORAGE OR MAXIMUM DEFICIT S (MEAN VALUES)

− $B)/\sigma$												Magnitude
.42	0.43	0.5	0.52	0.53	0.55	0.6	0.63	0.66	0.68	0.7	0.88	
HIRTY-EIGHT PHENOMENA; N = 96 YEARS)												
.34 .47	0.32 −0.52	....	0.24 −0.64	0.28 −0.58	0.24 −0.62	....	0.23 −0.67	0.20 −0.73	0.20 −0.76	....	0.09 −0.96	S/R log (S/R)
EVEN PHENOMENA; N = 430 YEARS)												
... ...	....	0.21 −0.68	....	....	....	0.15 −0.82	....	....	....	0.12 −0.92	....	S/R log (S/R)

12. The Relation Between Draft and Storage for Natural Phenomena

The results of a number of computations of the storage (S) required to guarantee a minimum draft (B) less than the mean have been compiled. Table 12 gives the mean values of these detailed computations. Curves of the exponential and square root forms have been fitted to these mean values and the results are shown in Figs. 8 and 9. As far as closeness of fit is concerned, over the range of observations, there is no significant difference between one type of curve and the other. At some future time, it may perhaps be possible to decide that one or the other has some theoretical justification, but this has not so far been possible. The results of the curve fitting may be studied in Table 13. The curves that have been fitted to the means of all the observations are shown in Fig. 10.

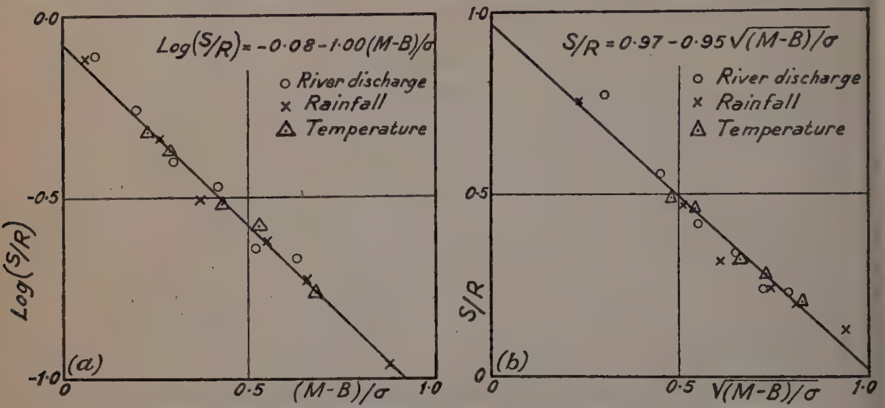

Fig. 8.—Analysis of River Discharge, Rainfall, and Temperature to Determine a Relation Between Draft B and Maximum Deficit S

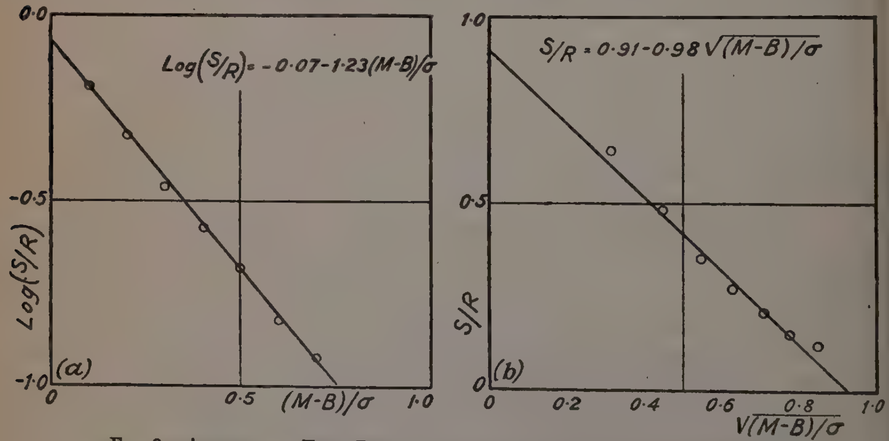

Fig. 9.—Analysis of Tree Rings, Varves, and Roda Gage to Determine a Relation Between Draft B and Maximum Deficit S

The equations show very little variation among the three groups of phenomena. For example, if $(M - B)/\sigma$ is 0.25, S/R from the exponential curve for each group is 0.47, 0.47, and 0.42 and from the square root curve for each group is 0.46, 0.49, and 0.42. These results, again, illustrate a striking uniformity.

TABLE 13.—COMPARISON OF CURVES SHOWING RELATION BETWEEN DRAFT *B* AND STORAGE OR MAXIMUM DEFICIT *S*

Fig. No.	Media	Log (S/R) =	S/R =
7	Cards, coins, and bonds (Eqs. 14)....	$-0.11 - 0.88\,(M - B)/\sigma$	$0.91 - 0.89\sqrt{(M - B)/\sigma}$
8	Rivers, rainfall, and temperatures....	$-0.08 - 1.00\,(M - B)/\sigma$	$0.97 - 0.95\sqrt{(M - B)/\sigma}$
9	Tree rings, varves, and Roda gage....	$-0.07 - 1.23\,(M - B)/\sigma$	$0.91 - 0.98\sqrt{(M - B)/\sigma}$
....	Mean of all observations............	$-0.08 - 1.05\,(M - B)/\sigma$	$0.94 - 0.96\sqrt{(M - B)/\sigma}$

Considering the curves based on the means of all the observations, which are given in Fig. 10, it will be seen that in spite of the difference in form, the foregoing expressions give practically identical curves over nearly all the range.

The important result shown by all the cases is that a small reduction in the guaranteed draft from the maximum value (the mean) makes a much greater proportional reduction in the storage required to maintain it. For example, a reduction of 0.1 σ diminishes the storage by about 35%. This fact provides a very practical method of applying a factor of safety to compensate for the natural variations of *R*.

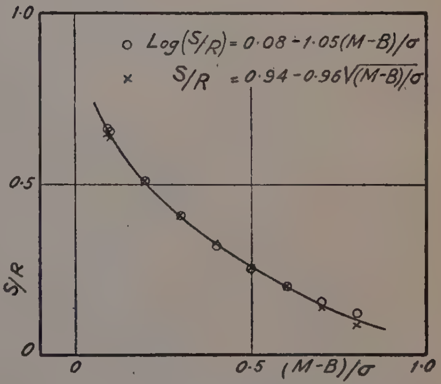

FIG. 10.—MEAN OF ALL OBSERVATIONS ANALYZED TO DETERMINE A RELATION BETWEEN DRAFT *B* AND MAXIMUM DEFICIT *S*

13. COMBINATION OF STORAGE FROM DIFFERENT SOURCES

The previous theory has dealt with the storage of one variable supply, as for example, the water flowing out of Lake Albert. It was shown (see Eq. 13) that, on the average, the capacity required to produce a steady discharge equal to the mean over a period of N years, on a stream whose standard deviation of annual discharge is σ, is given by $R = \sigma\left(\frac{N}{2}\right)^{0.72}$. If two streams are involved, with variations whose discharges have a normal distribution, then if σ_1 and σ_2 are their standard deviations, and σ is the standard deviation of their combined flow, it can be shown that

$$\sigma^2 = \sigma^2_1 + \sigma^2_2 + 2\,r_{1,2}\,\sigma_1\,\sigma_2 \qquad (15)$$

in which $r_{1,2}$ is the coefficient of correlation between the two discharges. It follows from Eq. 15 that, for a given value of N,

$$R^2 = R^2_1 + R^2_2 + 2\,r_{1,2}\,R_1\,R_2 \qquad (16)$$

If the two tributaries vary exactly together, $r_{1,2} = 1$ and $R = R_1 + R_2$. Except in this case the necessary capacity to equalize the discharge of the main stream is less when it is all in one reservoir on the main stream than when it is in reservoirs on each of the tributaries. It is possible under some circumstances for the discharge of the main stream to be equalized by a reservoir on a tributary.

14. Practical Applications

As mentioned in Section 1 the investigation arose out of the need to control the flow of the White Nile by reservoirs in Lake Albert and Lake Victoria. The scheme favored by Egypt was to construct a single large reservoir in Lake Albert of such a capacity as to equalize the outflow over a century. Such a scheme requires the simplest type of regulation, but it was not acceptable to Uganda because of the large area of land which would be flooded for long periods. The Uganda Government proposed the combination of a large storage reservoir in Lake Victoria with a small regulating reservoir in Lake Albert. Later a regulator at the exit of Lake Kioga in Uganda was seen to be a necessary part of this combination to eliminate the very variable effect of Lake Kioga on the Victoria Nile. The regulation of the combination is more complicated and has been the subject of much study.

In the more simple plan for Lake Albert the capacity of the reservoir would be large compared with the natural fluctuations of the lake. The method of working would have been to build the reservoir 15 years or 20 years before it was required to be in complete action, and, during this time, to fill it partly full, leaving a sufficient reserve capacity to control unusual floods that might occur after the project was working to full capacity and which would then do damage to the Sudan. The simplest working of the reservoir would be to send down a constant annual discharge slightly less than the mean, distributing this discharge as required through the year. This operation would lead to occasional years in which a discharge of more than the average would be necessary to avoid overfilling the reservoir.

This procedure was adopted in the report[2] for making estimates of the value of the projects and of the size of the works required. In these estimates the guaranteed discharge was less than the mean discharge by twice the probable error of the mean, plus a further small reduction. The actual procedure would depend on circumstances, beginning, no doubt, with what was thought to be a safe draft. If the reservoirs showed signs of reaching high levels, the draft would be increased and, if the contents diminished considerably, the draft would be reduced. During the early years of regulation, before the full draft was required, experience would be gained which would enable a sound scheme of regulation to be drawn up based on more information than is now available.

A modification of this simple regulation would include what the writer has called "virtual storage," in which an excess on the Ethiopian tributaries can be virtually stored by reducing the outflow from Lake Albert.

Lake Victoria, as the main storage reservoir, has the advantage that its area is 12.5 times that of Lake Albert. Therefore, a much lower dam will give the required storage; but the regulation is complicated by the fact that the natural range of rise and fall of the lake is greater than the extra range required for regulation. This would not matter if there were plenty of room for regulation, but the interests of the inhabitants around the lake place a definite limit on the height to which the lake can be raised, and the natural range of the lake plus the extra effect of regulation must be fitted within the allowable limits. This requirement fixes the capacity needed in Lake Albert, where extra storage over that required for the annual regulation is necessary to provide against unusual floods. The regulation of the two lakes in combination, therefore, departs from the simple one of a steady annual discharge out of Lake Victoria and a steady annual discharge out of Lake Albert, which would be almost automatic. The more complicated regulation will depend for success upon a close study of rainfall, runoff, discharges of tributaries, and trends of lake levels. However, the time that will elapse before the project is working will provide additional data, and experience in regulation.

It is fortunate that the Nile supply comes from two sources—Ethiopia and the Great African Lakes—and that, as far as available information goes, the correlation between their contributions is small. It is unlikely, therefore, that both sources would suffer extended periods of drought together, and this can provide a factor of safety against an unusual sequence of low years on the Lake Plateau which includes Uganda and parts of Kenya, Tanganyika, and the Belgian Congo. To take advantage of this the scheme of projects reported[2] must be considered as a whole and worked as a combination. Should an unusual succession of low years on the Lake Plateau cause a serious decline in the stored water, it would be necessary to make full use of virtual storage. Then, by beginning to store a little earlier in the main Nile reservoirs, the quota from the lakes could be reduced and the possibility of a dangerous depletion of the storage would be removed.

15. CONCLUSION

This paper has been mainly confined to theoretical matters, and the practical outcome therefrom has been reported elsewhere.[2]

16. ACKNOWLEDGMENTS

W. Borgquist and H. Munding, of the Kunglige Vattenfalls Styrelsen, Stockholm, Sweden; and C. E. P. Brooks, of the British Meteorological Office, supplied the writer with useful series of data on the thickness of annual deposits of clay in lake beds and other data (see Section 7). Basic data on the annual growth of American trees was obtained from the researches of A. E. Douglass.[10] The source of much of the meteorological information is the "World Weather Records" in the Smithsonian Miscellaneous Collections.[11] Other information is from various sources.[12]

Thanks are due to R. P. Black, director of research, Inspectorate General for Nile Control, for continued assistance and advice, and to W. Lawrence Balls, late of the Egyptian Ministry of Agriculture, for valuable suggestions. Nagib Boulos, Sayed Abdel Moneim Hassanein, and other members of the writer's staff rendered valuable assistance with a large amount of numerical work.

---

[10] "Climatic Cycles and Tree Growth: A Study of the Annual Rings of Trees in Relation to Climate and Solar Activity," by A. E. Douglass, *Publication No. 2890*, Carnegie Inst., Washington, D. C., 1928.

[11] "World Weather Records," collected from official sources by F. Exner, G. C. Simpson, G. Walker, H. H. Clayton, and R. C. Mossman, *Smithsonian Miscellaneous Collections*, Vol. 79, Smithsonian Inst., Washington, D. C., 1927.

[12] "Geochronologia Suecica Principles," by Gerard de Geer, Almqvist & Wiksells, Stockholm, Sweden, 1940.

APPENDIX. NOTATION

The following letter symbols have been adopted for use in this paper and in the discussion:

B = annual draft from storage S on a stream, less than the mean annual flow of the stream;
${}_mC_r$ = number of combinations of m different things taken r at a time;
d = departure:
  $\bar{d}$ = average difference between the range of the continued sums of 2 m N tosses of a single coin and N tosses of 2 m coins;
  Σd_e = continued sums, or cumulative totals;
e = base of natural logarithms;
G = number of gains;
h = total number of instances in which the losses exceed the gains;
h = number of heads r in N throws; $h = r\,N$;
K = an index $= \dfrac{R/\sigma}{\log N/2}$;
L = number of losses;
M = mean annual draft of water from storage;
m = number of coins in any toss N; ${}_mC_r$ (see C);
N = number of annual discharges recorded; period of observation; also, number of tosses, of a set of m coins; number of departures from a mean;
$n = N\,m$;
Q = discharge, annual;
R = range, from maximum to minimum, of the curve obtained by plotting the cumulative totals of the departures from the mean Σd_e, of the annual discharges of a stream, taken in order;
r = number of heads in m tosses of a coin; also, number of heads; ${}_mC_r$ (see C); $r_{1.2}$ = coefficient of correlation between two variables;
S = storage required to maintain a draft B less than the mean M; maximum storage deficit;
t = number of tails, $m - r$ in m throws or tosses;
x = abscissas;
y = ordinates;
Σ = "sum of"; and
σ = standard deviation:
  $\sigma_r = \sigma$ of the number of heads or tails;
  $\sigma_d = \sigma$ of the number of heads minus the number of tails—that is, $2\,r - 2\,m$.